

Global seismic tomography and modern parallel computers

Gaia Soldati (1), Lapo Boschi (2), Antonio Piersanti (1)

(1) *Istituto Nazionale di Geofisica e Vulcanologia, Roma, Italy*

(2) *E.T.H. Zürich, Switzerland*

Abstract

A fast technological progress is providing seismic tomographers with computers of rapidly increasing speed and RAM, that are not always properly taken advantage of. Large computers with both shared-memory and distributed-memory architectures have made it possible to approach the tomographic inverse problem more accurately. For example, resolution can be quantified from the resolution matrix rather than checkerboard tests; the covariance matrix can be calculated to evaluate the propagation of errors from data to model parameters; the L-curve method can be applied to determine a range of acceptable regularization schemes. We show how these exercises can be implemented efficiently on different hardware architectures.

keywords

Numerical inverse theory; seismology; global tomography; seismic resolution; Earth's mantle.

1 Introduction

Earth tomography requires the solution of inherently large, mixed-determined inverse problems. Since its very beginning, it has always involved the implementation of efficient algorithms on state-of-the-art computers.

In 1984, John Woodhouse and Adam Dziewonski published in J.G.R. one of the few articles that defined global seismic tomography. In their conclusions, they noted: "The calculations reported upon here were performed using an array processor (Floating Point, 120B) which, programmed in Fortran, is typically 10 times as fast as, say, a VAX 11/780. The path by path inversions and source determinations, using mantle waves and body waves, occupied the machine for approximately 60 hours, and each global iteration took approximately 7 hours. The inclusion of more data, and the extension of the method to higher frequencies, will probably require the use of the most advanced 'mainframe' computers." They are referring to an image of the Earth's upper mantle parameterized in terms of a cubic polynomial (vertically), and spherical harmonics up to degree 8 (laterally); that is to say, entirely specified by just 324 parameters (Woodhouse and Dziewonski 1984). The lower mantle was mapped in a separate inversion, as a linear combination of spherical harmonics up to degree 6, multiplied by Legendre polynomials up to degree 4, resulting in 245 parameters; "the size of the array needed to store the lower (or upper) triangle of the corresponding inner product matrix is 31,035; just a little less than the data memory of our AP-120B array processor, without which this study would not be feasible" (Dziewonski 1984).

In the following years, tomographers took advantage of a fast technological progress. The following generation of models published by the Harvard group covered the entire mantle, and were linear combinations of Chebyshev polynomials and spherical harmonics up to degree 12 (e.g., Su et al. 1994).

Other authors preferred a different approach, parameterizing the Earth's mantle with grids of voxels (Hager and Clayton 1989, Inoue et al. 1990). The "voxel" approach involved a substantially larger number of model coefficients ($\sim 10^4$), but also implied that the integral of data sensitivity multiplied by the basis functions be most often 0. The latter circumstance has important consequences, as we shall briefly illustrate: the tomographic linear inverse

problem is typically written

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{d}, \quad (1)$$

where the entries of the vector \mathbf{x} are the coefficients of the solution model (initially unknown), \mathbf{d} are the data (e.g., travel times), and \mathbf{A} is a matrix whose ij entry equals the integral, over the entire volume of the mantle, of the sensitivity of the i -th measurement to the Earth property (typically a seismic velocity or slowness) to be mapped, times the j -th basis function used to describe such property.

In the ray theory approximation, and if (as it will always be the case here) \mathbf{d} are observations of travel time anomaly, the volume integral reduces to an integral along the seismic ray path (sensitivity is 0 everywhere but on the ray path), and

$$A_{ij} = \int_{i\text{-th path}} f_j(\mathbf{r}(s)) ds, \quad (2)$$

where s denotes the incremental length along the ray path, identified by the equation $\mathbf{r} = \mathbf{r}(s)$, with \mathbf{r} denoting position; the N basis functions f_j are used to describe the slowness $\delta p(\mathbf{r}) = \sum_{k=1}^N x_k f_k(\mathbf{r})$ of the seismic phase in question. (It would be equivalent to formulate the problem in terms of velocity, but slowness happens to make algebra simpler.)

If the functions f_j are spherical harmonics, nonzero over the entire surface of the Earth, the integral in eq. (2) will be nonzero for all values of j . In the case of a voxel (spline, wavelet or other "local" functions) parameterization, the same integral will be 0, except for values of j whose corresponding voxel is crossed by a ray path. As anticipated, the matrix \mathbf{A} will therefore be dense if the model is parameterized in terms of spherical harmonics or other "global" functions; sparse if local functions are used.

The same is consequently true of the matrix $\mathbf{A}^T \cdot \mathbf{A}$, whose inverse has to be calculated for the least squares solution \mathbf{x}_{LS} to (1) to be found (a necessary step, as (1) in global seismology is strongly mixed-determined and does not have an exact solution),

$$\mathbf{x}_{LS} = (\mathbf{A}^T \cdot \mathbf{A} + \mathbf{D})^{-1} \cdot \mathbf{A}^T \cdot \mathbf{d}, \quad (3)$$

where the matrix D depends on the regularization scheme (e.g., Boschi and Dziewonski, 1999).

When A and $A^T \cdot A$ are sparse, equation (3) is most efficiently implemented via an iterative algorithm like CG or LSQR (e.g., Trefethen and Bau 1997). When they are dense, iterative algorithms become as slow as direct ones: the most efficient approach is then to implement (3) via Cholesky factorization of $A^T \cdot A$, and subsequent backsubstitution (e.g., Press et al. 1994, Trefethen and Bau 1997).

Implementation of LSQR does not require $A^T \cdot A$ to be calculated, as LSQR operates directly on A . A is bigger but sparser than $A^T \cdot A$, and sparse matrices can be stored efficiently (e.g., Press et al. 1994) to minimize the required disc space or RAM; A is therefore often less cumbersome than $A^T \cdot A$: it is so, at least, when global body wave travel time databases are inverted to derive global Earth structure. This, and the remarkable speed of LSQR in a regime of sparse A , allowed Grand (1994) and van der Hilst et al. (1997) to parameterize the Earth's mantle in terms of as many as $N \sim 250,000$ voxels: a three orders of magnitude increase in nominal resolution, with respect to the early studies of Dziewonski (1984) and Woodhouse and Dziewonski (1984) mentioned above.

Like Woodhouse and Dziewonski some ten years before, Grand and van der Hilst were exploiting available computers to their limit. Although sparse, A was still too large a matrix to be entirely fit on the RAM of a processor; LSQR, however, required only parts of A to be available at one time in the RAM: they could be run without ever storing A entirely in memory, at the expense of massive input from disc at each iteration. Given the number of solution coefficients, at least $\sim 10^2$ iterations were probably needed for LSQR to converge. This made even LSQR a very slow process, and left researchers with relatively little freedom to test the model resolution and the effect on the solution of different regularization schemes.

In view of the exponential growth in CPU speed over the last decade (e.g., Bunge and Tromp 2003), and the concurrent decrease in the price of

RAM, the current generation of global seismic tomographers has the means to approach the discipline in an entirely new fashion. With a fast processor, and enough RAM to store A entirely, not only LSQR is sped up enormously, but more time-consuming direct algorithms like Cholesky factorization of $A^T \cdot A$ also become feasible.

2 Cholesky factorization on a multiprocessor, shared-memory computer

It was originally proved by Paige and Saunders (1982), and later confirmed by Nolet (1985) and Boschi and Dziewonski (1999), with applications to mixed-determined tomographic problems, that LSQR converges correctly to the damped least squares solution (3), typically after a number of iterations $\ll N$. If A is sparse and sufficient RAM is available to store it, LSQR is therefore the most efficient algorithm to solve an inverse problem in the least squares sense. On the other hand, because it by-passes the calculation of $A^T \cdot A$ and the direct implementation of (3), LSQR cannot provide any measure of goodness of resolution and covariance, except by means of resolution, or "checkerboard" tests. The unreliability of the measure of resolution that those tests provide has been pointed out, for example, by L ev eque et al. (1993), and there have been efforts to derive the resolution matrix via an iterative, LSQR-type calculation (Zhang and McMechan, 1995; Minkoff, 1996; Nolet et al., 1999, 2001; Yao et al., 1999, 2001; Vasco et al., 2003).

The resolution matrix R can be thought of as the operator that relates "output" and "input" model in any checkerboard test; Menke (1989) shows that

$$R = (A^T \cdot A + D)^{-1} \cdot A^T \cdot A. \quad (4)$$

Clearly, R does not depend on the input model, and its similarity to the identity matrix is a measure of goodness of resolution. Its calculation requires that $A^T \cdot A + D$ be explicitly inverted, and this is most efficiently achieved

by Cholesky factorization of this matrix. Once the damped inverse of $\mathbf{A}^T \cdot \mathbf{A}$ is found, \mathbf{R} is quickly determined by backsubstitution, applied on the matrix $\mathbf{A}^T \cdot \mathbf{A}$ instead of the vector $\mathbf{A}^T \cdot \mathbf{d}$; this endeavour is not significantly more time-consuming than the implementation of (3) via Cholesky factorization and backsubstitution.

Boschi (2003) computed \mathbf{R} from the global teleseismic P-wave travel time database of Antolik et al. (2003), based upon the ISC Bulletins and including $\sim 600,000$ summary observations. He parameterized the Earth's mantle in terms of 20 vertical splines and 362 horizontal splines ($N = 7240$). Boschi's (2003) exercise was conducted on an IBM SP2 with 16 processors and 32 Gbytes of RAM. The IBM SP2 is a "shared-memory" machine: any processor can access at the same speed its entire RAM. This is a very useful feature when large matrices have to be factorized, a process that is inherently hard to parallelize; clusters of PCs are by construction "distributed-memory" computers, and hence more useful for the solution of forward, rather than inverse, problems.

Boschi (2003) notes that the most time-consuming step in deriving \mathbf{R} is the computation of $\mathbf{A}^T \cdot \mathbf{A}$, which took about twenty-four hours. This process was parallelized by subdividing the database in as many subsets as there were available processors, computing each subset's contribution to $\mathbf{A}^T \cdot \mathbf{A}$ on a separate processor, and eventually adding up the results. After computing $\mathbf{A}^T \cdot \mathbf{A}$, which needs to be done only once, \mathbf{x}_{LS} and \mathbf{R} can be derived in a few minutes; Boschi (2003) was thus able to perform numerous inversions, experimenting with the damping scheme and exploring the solution space, calculating each time the associated \mathbf{R} .

3 Running LSQR repeatedly on a distributed-memory cluster of PCs

The computer on which this article is being written, a Linux PC sitting on the second author's desk, is equipped with a dual processor and 3 GBytes of

RAM. One year ago, this much RAM costed just about 800 U.S. Dollars. We concluded section 1 pointing out that, as one can easily afford enough RAM to store arrays of $\sim 10^5$ elements, iterative algorithms like LSQR become extremely efficient. On this very computer, one LSQR inversion involving some 25,000 model coefficients runs to convergence in a matter of seconds.

Let us now show how, on a parallel, distributed-memory computer (a cluster of PCs), the resolution matrix \mathbf{R} can also be derived in a reasonable amount of time, without calculating and Cholesky-factorizing $\mathbf{A}^T \cdot \mathbf{A}$.

Implementing equation (4) is equivalent to implementing N times equation (3), replacing each time \mathbf{d} with a different column of \mathbf{A} (recall that N denotes the number of model parameters, and hence the number of columns of \mathbf{A}). \mathbf{R} can therefore be derived by means of N independent LSQR inversions of \mathbf{A} , without finding $\mathbf{A}^T \cdot \mathbf{A}$. When only one or few processors are available, and with $N \sim 10^5$ as in some of the experiments mentioned above, this process would be extremely time-consuming, to the point of not being worthwhile. If a relatively large parallel machine is available, however, the problem can be easily parallelized, by simply subdividing the N inversions into N/n_P subsets, n_P denoting the number of processors. Each subset of inversions is then performed independently on a separate processor, and the time needed to compute \mathbf{R} is reduced by a factor n_P .

It should be noted that the most time-consuming step, input of \mathbf{A} from disc to RAM, needs to be performed only once per processor, no matter how many inversions are then run on each processor.

Figures 1 and 2 illustrate how this procedure applies to a real inverse problem. We describe the distribution of P-velocity heterogeneities in the Earth's mantle in terms of a grid of voxels of constant horizontal extent; voxel functions guarantee that \mathbf{A} be more sparse than in the case of splines. We invert, again, the P-wave travel time database of Antolik et al. (2003). Following, e.g., Inoue et al. (1990, section 3.3.1 and figure 2), we select roughness minimization as our only regularization criterion, and perform a number of preliminary inversions, at different parameterization levels, to assess the

dependence of the solution on the regularization parameter. Plotting misfit to the data (defined as $1 - \text{the variance reduction}$) against "total roughness" (the integral of the surface gradient of the model over the entire solid angle is computed for each layer of the model, and then the RMS is taken) in figure 1, we find, for each parameterization, a set of points aligned along the expected L-shaped curve (e.g., Hansen 1992; Boschi et al., 2006). Each point on the L-curve corresponds to a model derived at this preliminary stage, and the roughness damping parameter grows monotonically with increasing misfit. The shape of the curve, resembling the letter L, confirms that the data contain coherent and statistically significant information; the decrease in misfit is very fast in an overdamped regime, where a small reduction in the regularization parameter, and therefore a small increase in model complexity, is sufficient to improve the data fit substantially. The white noise that the data necessarily contain, and that regularization is supposed to eliminate, is harder to fit, even with large increases in model complexity: this is why the curve tends to become horizontal in the right part of the plot. Solution models lying in the vertical and horizontal portions of the L-curve can be discarded as overdamped and underdamped, respectively; preferred models should be chosen near its corner.

The selection of a damping scheme has always been a largely arbitrary process in global seismic tomography. The L-curve criterion is a way to reduce this arbitrariness. It is practical so long as a large number of LSQR inversions can be performed in a short time, and we have seen how this is made possible by simultaneous storage of the entire matrix \mathbf{A} in memory, and/or availability of multiple processors.

After so selecting optimal roughness damping parameters at all parameterization levels, we restrict ourselves to the case of 5° voxels. We show in figure 2 the corresponding resolution matrix \mathbf{R} as derived with multiple runs of LSQR. As to be expected (Boschi, 2003), \mathbf{R} is quite different from the identity matrix; entries smaller than 1 on its diagonal indicate that the amplitude of velocity heterogeneities in the corresponding voxel is underesti-

mated. Entries different from zero away from the diagonal identify episodes of fictitious coupling between model coefficients; naturally, the value of R_{ij} is proportional to the amount of coupling ("trade-off", "smearing"...) between the i -th and j -th voxels (entries x_i and x_j of the solution vector).

4 Performance and accuracy of direct vs. iterative implementations

In analogy with Yao et al. (1999), we calculate R associated with one given database and one choice of parameterization and regularization, both in the direct (Cholesky, section 2 above) and iterative (LSQR, section 3) approaches, and compare the results. As opposed to singular value decomposition (SVD), the direct algorithm implemented by Yao et al. (1999), Cholesky factorization does not involve the cancellation of the smallest singular factors (Press et al., 1994), so that in our experiment regularization is entirely controlled by the matrix D , and is therefore exactly equivalent in the direct and iterative calculations.

We implement equation (4) exactly, by Cholesky factorization of $A^T \cdot A + D$, for the 5° -voxel parameterization described in section 3, and applying the same regularization scheme that lead to R in figure 2. The result is shown in figure 3, and in figure 4 two lines of the directly- and iteratively-calculated R s are compared in a geographic view. Differences are everywhere small, and negligible for i, j such that $R_{ij} > 0.2$. Discrepancies between R calculated iteratively and directly (with SVD) by Yao et al. (1999, figure 5) appear to be larger.

R in figures 2 through 5 is a $24,840 \times 24,840$ matrix, as opposed to the $7,240 \times 7,240$ R of Boschi (2003). With 24,840 free parameters, Cholesky factorization, backsubstitution (via the Lapack routines SPOTRF and SPOTRS, respectively), and all necessary input/output from and to disc take about 10 hours on a shared-memory Compaq "Alpha" computer (an ES45 with 10Gbytes RAM and 4 CPUs at 1250MHz); To compare this performance

with that of repeated LSQR on a PC-cluster, it should be kept in mind that, in the latter architecture, computation time scales perfectly with the number of processors; one LSQR inversion with 24, 840 free parameters, and applying Paige and Saunders' (1982) criterion to evaluate convergence, currently takes ~ 1 minute on a standard PC.

5 The covariance matrix

\mathbf{R} describes the fictitious coupling between solution coefficients (model parameters); it depends on the geographic distribution of sources and stations, and on the shape of ray paths, but not on the quality of inverted observations. The covariance of solution coefficients depends, instead, on the error and covariance of the initial data, and on the error amplification occurring in the inversion (Menke, 1989, section 3.11). In the assumption that seismic data be uncorrelated and all have equal variance σ^2 , Menke (1989, equation 3.48) introduces a covariance matrix

$$\mathbf{C} = \sigma^2 (\mathbf{A}^T \cdot \mathbf{A} + \mathbf{D})^{-1} \cdot \mathbf{A}^T \cdot \left[(\mathbf{A}^T \cdot \mathbf{A} + \mathbf{D})^{-1} \cdot \mathbf{A}^T \right]^T \quad (5)$$

(the regularization matrix \mathbf{D} was not included explicitly in Menke's (1989) formula). Equation (5) can be rewritten

$$\mathbf{C} = \sigma^2 (\mathbf{A}^T \cdot \mathbf{A} + \mathbf{D})^{-1} \cdot \mathbf{A}^T \cdot \mathbf{A} \cdot \left[(\mathbf{A}^T \cdot \mathbf{A} + \mathbf{D})^{-1} \right]^T, \quad (6)$$

and making use of (4)

$$\mathbf{C} = \sigma^2 \mathbf{R} \cdot \left[(\mathbf{A}^T \cdot \mathbf{A} + \mathbf{D})^{-1} \right]^T. \quad (7)$$

After Cholesky factorizing $\mathbf{A}^T \cdot \mathbf{A} + \mathbf{D}$, we find \mathbf{C} by (i) backsubstitution of the $N \times N$ identity matrix, and (ii) dot-product (via the Lapack routine SGEMM) of the (transposed) result with \mathbf{R} . After \mathbf{R} is read from disc or calculated again, the process takes about 10 more hours, with $N = 24, 840$, on the shared-memory machine described in section 4 above. Figure 6 shows \mathbf{C} , derived in the same parameterization (5° voxels) and regularization as

figures 2 through 5 above, and assuming for Antolik et al.’s (2003) database a standard deviation $\sigma = 0.5$ s (Antolik, personal communication, 2005).

We have not found an effective approach to calculating C on a distributed-memory cluster.

Except for the upper mantle, where the solution is less stable owing to strong nonuniformities in the data coverage, C (figure 6) is relatively close to diagonal, indicating that errors in model parameters are not strongly correlated. The N diagonal entries of C can be interpreted as squared errors associated with the corresponding model parameters (Menke, 1989; Tarantola, 2005): after taking their square root and multiplying it by 100 (in a voxel parameterization, solution coefficients coincide with velocity heterogeneities in the corresponding voxel, which are typically expressed in percent), we show in figure 7 each diagonal entry of C at the corresponding voxel. As to be expected, error is smallest in regions of good data coverage, e.g. the upper and mid-mantle underlying North America and Eurasia, where seismic stations are most densely distributed; it is highest at the top of the upper mantle, where the almost vertical geometry of teleseismic ray paths poses a significant limit to resolution (hence strong “smearing”); it grows with increasing depth in the bottom layers of the lower mantle, sampled more uniformly than shallowest regions, but by a decreasing number of ray paths. Mapped P-velocity anomalies from the observations considered here (e.g., Boschi & Dziewonski, 1999; Boschi, 2003) range between $\pm 1\%$ in most of the mantle, so that the error of $\pm 0.15\%$ or less that we have derived from C is generally nonnegligible, but small.

6 Summary

We have presented two approaches to the solution of large mixed-determined inverse problems, both exploiting the quickly increasing speed and RAM of modern computers (e.g., Bunge and Tromp 2003). We have verified (section 4) that the two approaches, applied to the same problem, yield coincident

results.

The first approach, described in section 2, is inherently sequential, and is best applied to shared-memory computers. It rests on the direct implementation of the least squares formula to derive the least squares solution x_{LS} and model resolution matrix \mathbf{R} associated with the inverse problem $\mathbf{A} \cdot \mathbf{x} = \mathbf{d}$. x_{LS} and \mathbf{R} are computed with one Cholesky factorization of $\mathbf{A}^T \cdot \mathbf{A}$ and $N + 1$ repetitions of the backsubstitution process, N being the number of model coefficients. In this approach, the derivation of \mathbf{R} is thus relatively fast. Unfortunately, the number of floating point operations required to Cholesky-factorize $\mathbf{A}^T \cdot \mathbf{A}$ grows like N^3 , as finer parameterizations are implemented (e.g., Trefethen and Bau 1997, page 175). Likewise, as N grows, the size of $\mathbf{A}^T \cdot \mathbf{A}$ grows like N^2 , and RAM can also become an issue: routines performing Cholesky factorization, available in the literature (e.g., Press et al. 1994) or through optimized libraries, do not allow for efficient storage of $\mathbf{A}^T \cdot \mathbf{A}$ (which could be quite sparse), and require a comparably large additional amount of RAM to be left free for temporary storage.

The second approach (section 3) involves the repeated application of an iterative, CG-type algorithm (LSQR in our implementation). x_{LS} is found after one run of LSQR, the calculation of \mathbf{R} requires N runs of the same algorithm. However, we have shown that the problem can be simply parallelized, and it is thus most appropriate for implementation on distributed-memory PC-clusters. As N grows, the growth in the number of floating-point operations will not be as fast as in the case of the first approach, and it will be relatively cheap to speed up the process by simply making use of a few more processors. As long as \mathbf{A} is sparse, which is always the case in the ray-theory approximation and with local-basis-function (voxels, splines...) parameterizations, the amount of disc space (and/or RAM) needed to store \mathbf{A} also grows more slowly with increasing N than that needed to store $\mathbf{A}^T \cdot \mathbf{A}$: in the experiment discussed in section 3, \mathbf{A} occupies roughly 1 Gbyte of RAM or disc space; $\mathbf{A}^T \cdot \mathbf{A}$ would need twice this amount, plus the temporary storage mentioned above.

One drawback of the CG/LSQR multi-processor approach resides in the difficulty of computing the covariance matrix C . We have shown in section 5 how C can instead be computed via the first, “sequential” approach, and we have made use of C to evaluate model error in typical, global tomographic inversions of seismic travel time observations (figure 7).

In summary, both approaches should prove profitable, depending on the available hardware. The optimization of tomographic algorithms for use with modern computers is leading to a better understanding of the tomographic inverse problem, and to more reliable evaluations of model quality and resolution.

ACKNOWLEDGMENTS

Our research is part of SPICE (Seismic wave Propagation and Imaging in Complex media: a European network), a Marie Curie Research Training Network in the 6th Framework Program of the European Commission; we are grateful to the coordinators at LMU Munich. We thank Domenico Giardini for his support and encouragement. All figures were done with GMT (*Wessel and Smith 1991*).

References

- Antolik, M., Gu, Y. J., Ekström, G. & Dziewonski, A. M., 2003. J362D28: a new joint model of compressional and shear velocity in the mantle, *Geophys. J. Int.*, **153**, 443–466.
- Boschi, L., 2003. Measures of resolution in global body wave tomography, *Geophys. Res. Lett.*, **30**, NO. 19, 1978, doi:10.1029/2003GL018222.
- Boschi, L. & Dziewonski, A. M., 1999. “High” and “low” resolution images of the Earth’s mantle - Implications of different approaches to tomographic modeling, *J. geophys. Res.*, **104**, 25,567–25,594.

- Boschi, L., T. W. Becker, G. Soldati, and A. M. Dziewonski, 2006. On the relevance of Born theory in global seismic tomography, *Geophys. Res. Lett.*, **33**, L06302, doi:10.1029/2005GL025063.
- Bunge, H.-P. & Tromp, J., 2003. Supercomputing moves to universities and makes possible new ways to organize computational research, *EOS, Trans. Am. geophys. Un.*, **84**, 30–33.
- Dziewonski, A. M., 1984. Mapping the lower mantle: determination of lateral heterogeneity in P velocity up to degree and order 6, *J. geophys. Res.*, **89**, 5929–5952.
- Grand, S. P., 1994. Mantle shear structure beneath the Americas and surrounding oceans, *J. geophys. Res.*, **99**, 11,591–11,621.
- Hager, B. H., & R. W. Clayton, 1989. Constraints on the structure of mantle convection using seismic observation, flow models, and the geoid, in *Mantle Convection-Plate Tectonics and Global Dynamics*, edited by W. R. Peltier, pp. 657–763, Gordon and Breach, Newark, N. J.
- Hansen, P. C., 1992. Analysis of discrete ill-posed problems by means of the L-curve, *SIAM review*, **34**, 561–580.
- Inoue, H., Fukao, Y., Tanabe, K. & Y. Ogata, 1990. Whole mantle P wave travel time tomography, *Phys. Earth planet. Inter.*, **59**, 294–328.
- L ev eque, J. J., L. Rivera, & G. Wittlinger, 1993. On the use of the checkerboard test to assess the resolution of tomographic inversions, *Geophys. J. Int.*, **115**, 313–318.
- Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory*, rev. ed., Academic, San Diego.
- Minkoff, S. E., 1996. A computationally feasible approximate resolution matrix for seismic inverse problems, *Geophys. J. Int.*, **126**, 345–359.

- Nolet, G., 1985. Solving or resolving inadequate and noisy tomographic systems, *J. Comput. Phys.*, **61**, 463–482.
- Nolet, G., R. Montelli, & J. Virieux, 1999. Explicit, approximate expressions for the resolution and a posteriori covariance of massive tomographic systems, *Geophys. J. Int.*, **138**, 36–44.
- Nolet, G., R. Montelli, & J. Virieux, 2001. Reply to comment by Z. S. Yao, R. G. Roberts and A. Tryggvason on “Explicit, approximate expressions for the resolution and a posteriori covariance of massive tomographic systems”, *Geophys. J. Int.*, **145**, 315.
- Paige, C. C., & M. A. Saunders, 1982. LSQR: an algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Soft.*, **8**, 43–71.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & B. P. Flannery, 1994. *Numerical Recipes in FORTRAN*, Cambridge University Press, U. K.
- Soldati, G. & L. Boschi, 2004. Whole Earth tomographic models: a resolution analysis, *EOS, Trans. Am. geophys. Un.*, **85**(47), Fall Meet. Suppl.
- Su, W.-J., R. L. Woodward & A. M. Dziewonski, 1994. Degree-12 Model of Shear Velocity Heterogeneity in the Mantle, *J. geophys. Res.*, **99**, 4945–4980.
- Tarantola, A., 2005. *Inverse Problem Theory and Model Parameter Estimation*, SIAM, Philadelphia.
- Trefethen, L. N. & D. Bau III, 1997. *Numerical Linear Algebra*, Soc. for Ind. and Appl. Math., Philadelphia, Penn.
- van der Hilst, R. D., S. Widiyantoro & E. R. Engdahl, 1997. Evidence for deep mantle circulation from global tomography, *Nature*, **386**, 578–584.
- Vasco, D. W., L. R. Johnson & O. Marques, 2003. Resolution, uncertainty, and whole-Earth tomography, *J. geophys. Res.*, **108**, 2022, doi:10.1029/2001JB000412.

- Wessel, P. & W. H. F. Smith, 1991. Free software helps map and display data. *EOS, Trans. Am. geophys. Un.*, **72**, 445–446.
- Woodhouse J. H. & A. M. Dziewonski, 1984. Mapping the upper mantle: three-dimensional modeling of Earth structure by inversion of seismic waveforms, *J. geophys. Res.*, **89**, 5953–5986.
- Yao, Z. S., R. G. Roberts & A. Tryggvason, 1999. Calculating resolution and covariance matrices for seismic tomography with the LSQR method, *Geophys. J. Int.*, **138**, 886–894.
- Yao, Z. S., R. G. Roberts, & A. Tryggvason, 2001. Comment on “Explicit, approximate expressions for the resolution and a posteriori covariance of massive tomographic systems” by G. Nolet, R. Montelli and J. Virieux, *Geophys. J. Int.*, **145**, 307–314.
- Zhang, J. & G. A. McMechan, 1995. Estimation of resolution and covariance for large matrix inversions, *Geophys. J. Int.*, **121**, 409–426.

Figure 1: Data misfit achieved by a set of solution models, vs. the integrated roughness of each model. This measure of model complexity is normalized against model RMS. Least squares solutions were found from a wide range of values of the roughness minimization parameter, and no other minimization constraint. We repeated the experiment with voxels of lateral extent $15^\circ \times 15^\circ$, $10^\circ \times 10^\circ$, $7.5^\circ \times 7.5^\circ$, $6^\circ \times 6^\circ$, $5^\circ \times 5^\circ$, $3.75^\circ \times 3.75^\circ$, $3^\circ \times 3^\circ$, $2.5^\circ \times 2.5^\circ$, and constant vertical thickness (~ 200 km) (Soldati and Boschi, 2004): corresponding solutions align on different L-curves, and squares of decreasing size correspond to increasingly fine parameterization.

Figure 2: (Top) $24,840 \times 24,840$ (5° -voxel grid) resolution matrix \mathbf{R} for the chosen regularization scheme, averaged (Boschi, 2003) so that it can be plotted in this limited space; here, vertical tradeoffs are most evident. (Bottom) Zooms on \mathbf{R} , not averaged, at two selected layers (left: mid mantle at ~ 1300 km depth; right: lower mantle at ~ 2600 km).

Figure 3: Same as figure 2, but \mathbf{R} was computed by Cholesky factorization of $\mathbf{A}^T \cdot \mathbf{A} + \mathbf{D}$, as described in section 4.

Figure 4: Rows of \mathbf{R} (figures 2 and 3) associated with a relatively well resolved voxel i located in the mantle under Japan, at 700 km mean depth, from the parallel LSQR (left) and Cholesky (right) approaches. For each value of j , the color of the j -th voxel depends on the value of R_{ij} ; R_{ij} is a measure of fictitious trade-off between i -th and j -th model parameters.

Figure 5: Same as figure 4, but voxel i , less well resolved, is located under Central America, at 2200 km mean depth.

Figure 6: $24,840 \times 24,840$ (5° -voxel grid) covariance matrix \mathbf{C} associated with the same data, parameterization and regularization as \mathbf{R} above, defined as in section 5, and derived by Cholesky factorization and backsubstitution.

Figure 7: Absolute error on mapped percent P-velocity heterogeneity, calculated from the diagonal entries of \mathbf{C} (figure 6) and plotted at each corresponding model voxel. All 15, ~ 200 km thick layers of the 5° -voxel grid are shown; the shallowest layer is at the top and to the left, the plot below corresponds to the second shallowest layer, and so on; the deepest layer (~ 2700 km depth to core-mantle boundary) is at the bottom and to the right. Constant, uncorrelated variance $\sigma = 0.5$ is assumed on all travel-time observations.

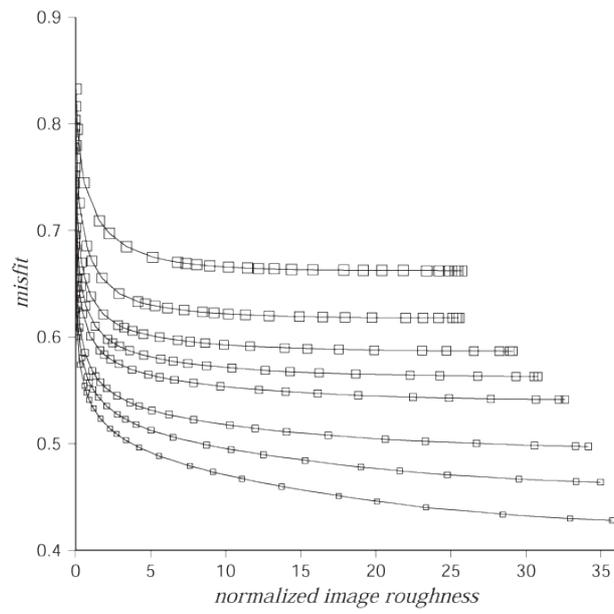


Figura 1

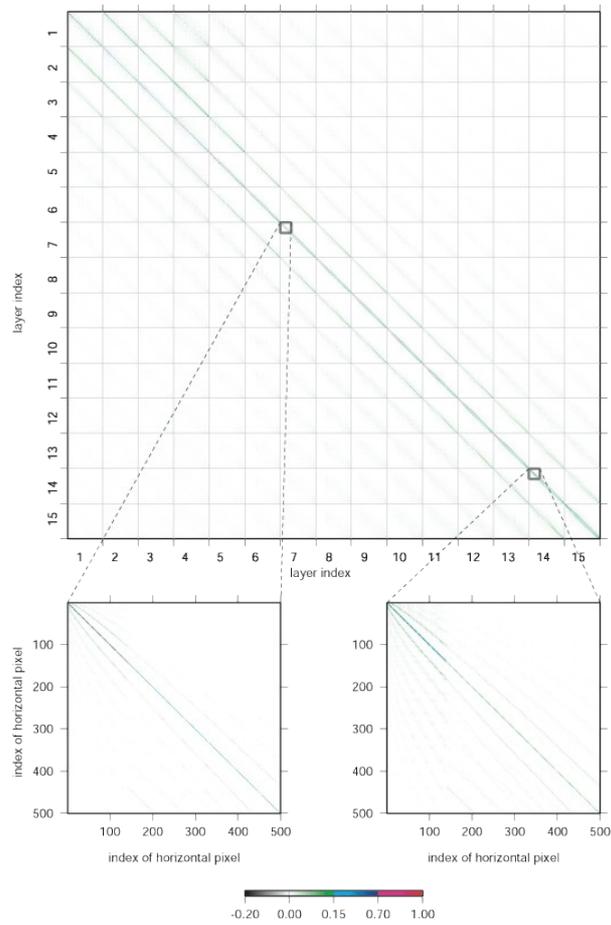


Figura 2

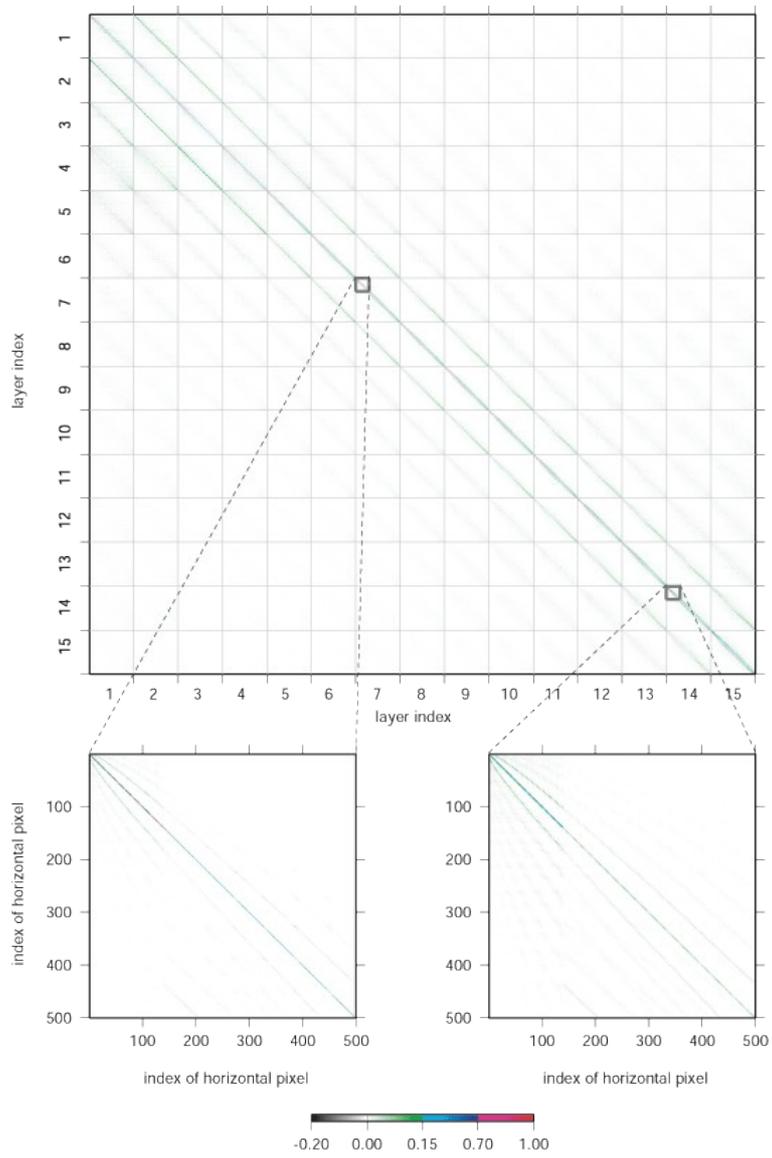


Figura 3

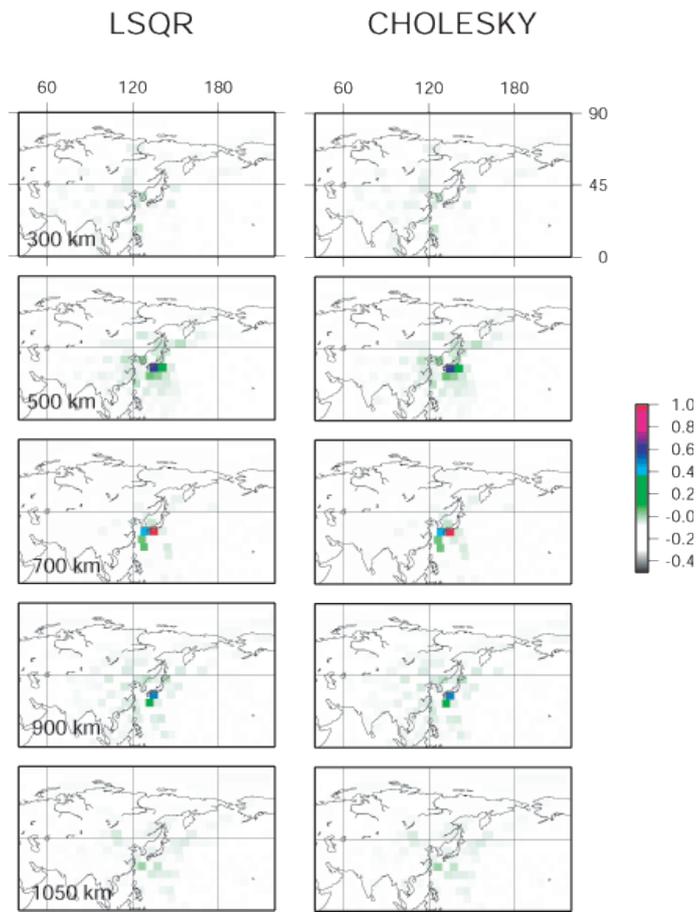


Figura 4

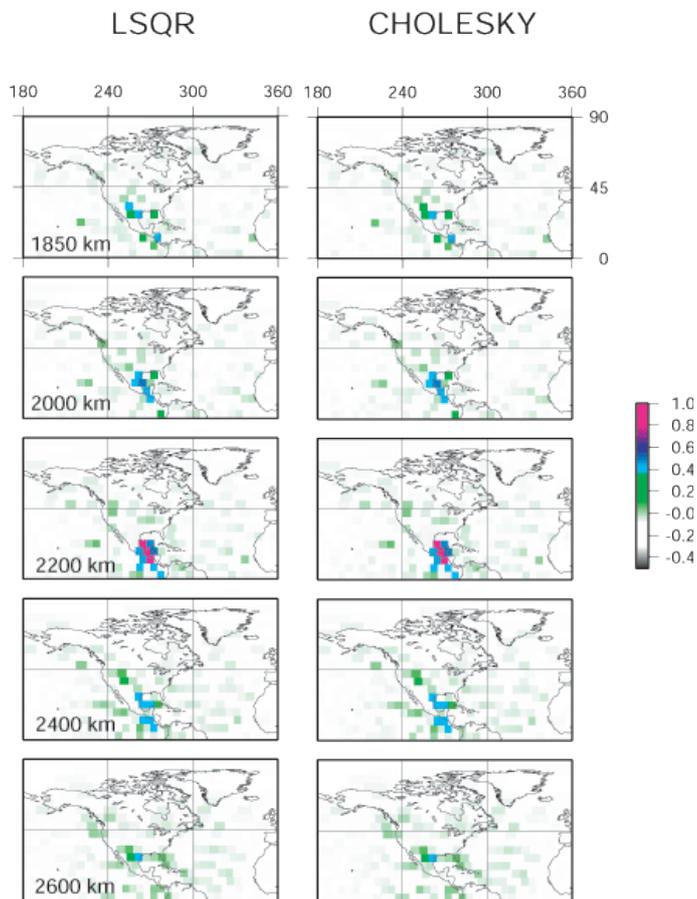


Figura 5

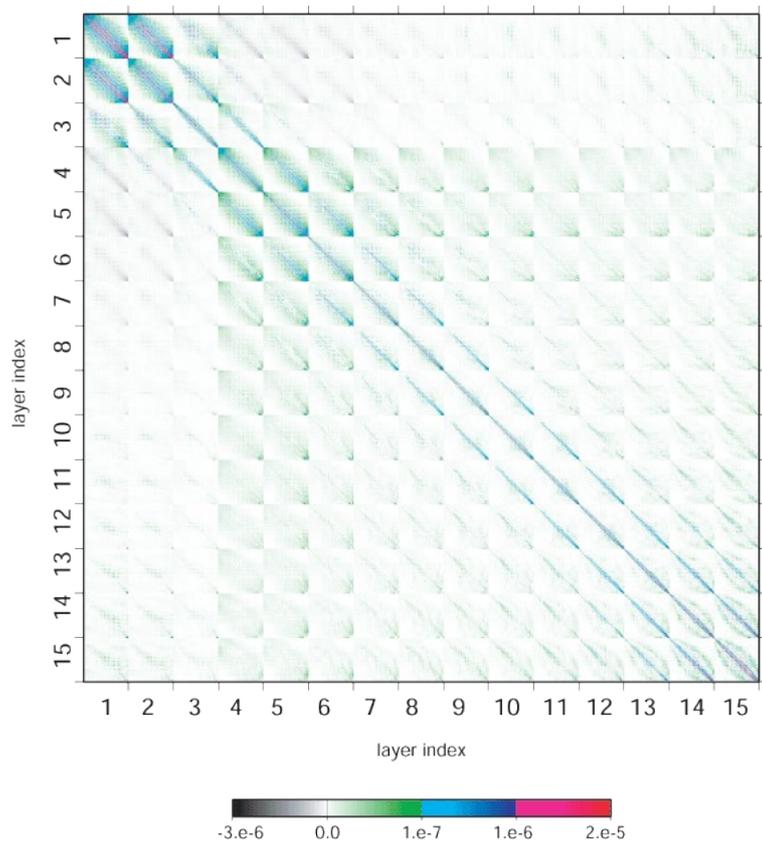


Figura 6

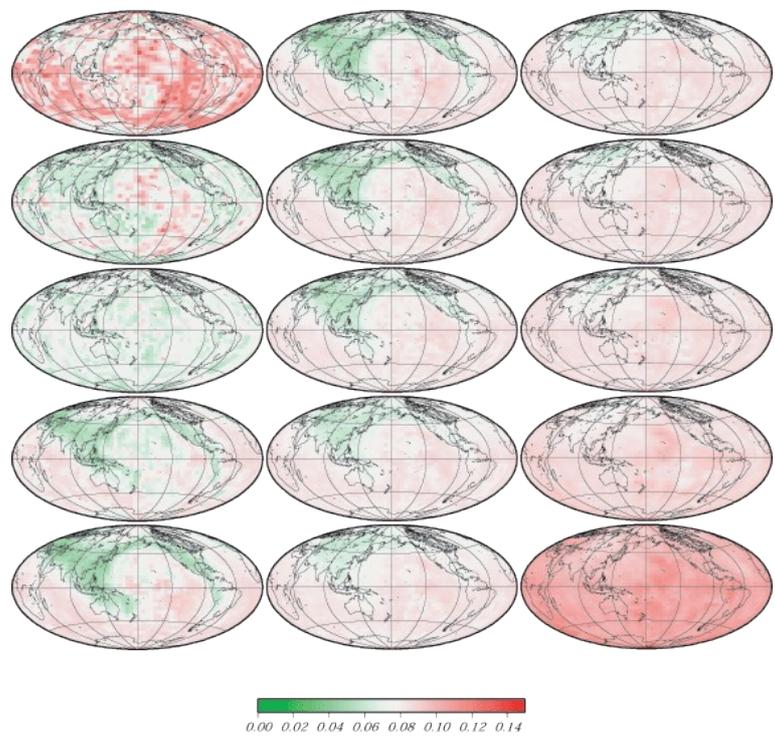


Figura 7